Identification and visualization of novel bacteria in activated sludge

Kristensen, Jannie Munk

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

IDENTIFICATION AND VISUALIZATION OF NOVEL BACTERIA IN ACTIVATED SLUDGE

BY
JANNIE MUNK KRISTENSEN

DISSERTATION SUBMITTED 2019

AALBORG UNIVERSITY
DENMARK

IDENTIFICATION AND VISUALIZATION OF NOVEL BACTERIA IN ACTIVATED SLUDGE

BY

Jannie Munk Kristensen

AALBORG UNIVERSITY
DENMARK

Dissertation submitted December 29, 2019

Dissertation submitted: December 29, 2019

PhD supervisor: Prof. Per Halkjær Nielsen,
Aalborg University

PhD committee: Associate Professor Peter Roslev (chairman)
Aalborg University

Professor Andreas Schramm
Aarhus University

Professor Malte Hermansson
University of Gothenburg

PhD Series: Faculty of Engineering and Science, Aalborg University

Department: Department of Chemistry and Bioscience

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-576-5

Published by:
Aalborg University Press
Langagervej 2
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2020

Dissertation submitted: December 29, 2019

PhD supervisor: Prof. Per Halkjær Nielsen,
Aalborg University

PhD committee: Associate Prof. Peter Roslev,
Aalborg University (Chairman)

Prof. Malte Hermansson,
University of Gothenburg, Sweden

Prof. Andreas Schramm,
Aarhus University

PhD Series: Faculty of Engineering and Science, Aalborg University

ISSN: xxxx- xxxx
ISBN: xxx-xx-xxxx-xxx-x

Published by:
Aalborg University Press
Skjernvej 4A, 2nd floor
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2019

ENGLISH SUMMARY

Microorganisms play a vital role in most processes essential for solving environmental challenges both now and in the future. Microbial processes in wastewater treatment systems are essential for resource recovery and the production of bioenergy. A prerequisite for overcoming environmental challenges is to identify the microorganisms performing the key processes in these systems and explore their functions. Development of DNA sequencing techniques has in recent years allowed the exploration of microbial diversity through amplicon sequencing of the 16S rRNA gene. Further development of molecular methods and bioinformatics has resulted in methods to retrieve full-length 16S rRNA genes at scale from complex samples. This has enabled the exploration of novelty through new more complete databases and the possibilities to design more specific primers and probes.

The aim of this project was to gain detailed knowledge of physiology and ecology for key species in the activated sludge process. The knowledge was collected to facilitate the long-term goal of biological process optimization and troubleshooting at wastewater treatment plants worldwide. This project is a central part of the Microbial Database of Activated Sludge (MiDAS) field guide efforts to collect the global knowledge in an open database. Key organisms were identified by creating and utilizing a full-length 16S rRNA gene database of all Danish wastewater treatment plants. In addition, a curated MiDAS taxonomy was developed, and combined with historical sampling of 1000+ samples for amplicon profiling of the v1-3 region of the 16S rRNA gene. The many samples and the 16S rRNA gene database were used to prioritize the efforts for designing molecular probe sets for visualization using Fluorescence *in situ* hybridization (FISH). To link identity with function, organisms were also explored through genomic studies and Raman spectroscopy.

We developed a pipeline to create an ecosystem specific database of 16S rRNA genes from activated sludge and utilized that database to design species specific probes. An analysis of existing probes for the 30 most abundant genera in wastewater treatment plants were made, revealing that only 19 of them were targeted by a probe, and many existing probes need to be reevaluated with the expanded databases. It was demonstrated that it is possible in some cases to design species specific FISH probes based on the 16S rRNA gene, in other cases it was not possible due to high similarity. A new probe set was designed for the ammonia oxidizing bacteria genera *Nitrosomonas* and *Nitrosospira*. The role in wastewater treatment plants of bacteria from the phylum Acidobacteria was investigated and some bacteria from this phylum were found to be putative PAOs. Pathogenic species of the genus *Arcobacter* was found to be incompletely removed by biological wastewater treatment thus leaving a high abundance in the effluent.

DANSK RESUME

Mikroorganismer har en vigtig rolle i forbindelse med rensning af spildevand, de omdanner næringsstoffer og sikrer human sundhed. Sammensætningen af bakterier i spildevandet, som udfører vigtige rensefunktioner, er meget kompleks og divers, og historisk har det været et langsommeligt arbejde at karakterisere disse. Det betyder at der i dag, på trods af mange års forskning, kun er begrænset viden om de bakterier, som er vigtige i forbindelse med rensning af spildevand. De senere års udvikling indenfor molekylære DNA sekventeringsmetoder har givet mulighed for at undersøge bakterier i deres eget miljø ved hjælp af deres 16S rRNA gen. Hvor gamle sekventeringsmetoder kun har givet mulighed for at læse korte stykker af dette gen, giver nye metoder nu muligheden for at undersøge hele sekvensen af 16S-genet fra bakterierne. Det har ændret de muligheder, der er for at undersøge mikrobiel diversitet, og sætte navne på også de ukendte bakterier gennem mere komplette 16S referencedatabaser. Dette giver også mulighed for at designe nye primere og FISH-prober, samt at revidere specificiteten af de gamle.

Formålet med dette projekt er at bibringe ny viden om hidtil ukendte bakterier i forbindelse med rensning af spildevand. Det er sket ved ved etablere en ny fuld-længde 16S database med MiDAS taxonomi, kombineret med V1-3 amplicon sekventering til at udforske det mikrobielle samfund i danske renseanlæg samt designe nye FISH-prober til visualisering af bakterier med nye specifikke probesæt. Udvalgte organismer blev desuden også udforsket ved hjælp af genomsekventering og Raman spektroskopi.

Vi har udviklet en metode til at etablere en økosystemspecifik database for aktivt slam, og brugt den database til at lave en analyse af hvilke FISH prober som findes til de mest forekommende bakterieslægter i renseanlæg. Kun 19 ud af de 30 havde en FISH-probe til at dække dem og mange af disse prober burde blive revideret nu der er kommet flere 16S rRNA gensekvenser i databaserne. Vi påviste at det nogle gange er muligt at designe FISH prober helt ned på artsniveau mens 16S rRNA gensekvensen for andre organismer er for ens til at differentiere dem.

Et nyt probesæt blev designet for de ammonium oxiderende bakterier, bakterieslægterne *Nitrosomonas* og *Nitrosospira*. Bakterier fra rækken Acidobacteria blev undersøgt og der er indikationer på at de kan være mulige fosforakkumulerende bakterier. Sygdomsfremkaldende arter af slægten *Arcobacter* blev fundet i en høj forekomst i renset spildevand, hvilket indikerer at de ikke fjernes effektivt af biologisk spildevandsbehandling.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor Per, for giving me this opportunity to join the scientific community as a part of the Center for Microbial Communities and for his exceptional guidance throughout this journey, for being patient, listening to my crazy ideas and believing in me. I have enjoyed your valuable inputs and enthusiastic support during the project.

I would also like to thank all my colleagues, for many great memories of conferences, parties and cakes. My office mates from which I have learned a lot from during the years and for making the office a cheerful place. To members of the MiDAS and FISH-groups for your hard work, and especially Lee-Ann who has been invaluable for my laboratory work. Thank you, Jakob and Martin, for your company at our many shared PhD courses and travels and for sharing this journey with me.

Thank you Petra Pjevac, Holger Daims and Michael Wagner for a warm welcome and a life changing collaboration during my research stay at DOME in Vienna, for teaching me FISH-probe design and microscopy, and to the entire DOME-group for teaching me how to enjoy the Viennese lifestyle as a true DOME.

And lastly thanks to my beloved family and friends who may not always understand what I do but always have supported me patiently and shown me that life could be other things than microbes under the microscope.

TABLE OF CONTENTS

LIST OF PAPERS

Paper 1 MS. Dueholm, KS. Andersen, F. Petriglieri, SJ. McIlroy, M. Nierychlo, J. Petersen, **JM. Kristensen**, E. Yashiro, SM. Karst, M. Albertsen, and PH. Nielsen. Comprehensive ecosystem-specific 16S rRNA gene databases with automated taxonomy assignment (AutoTax) provide species-level resolution in microbial ecology ***Preprint on BioRxiv***

Paper 2 M. Lukumbuzya, **JM. Kristensen**, P. Pjevac, K. Kitzinger, A. Pommerening-Röser, PH. Nielsen, H. Daims, M. Wagner. A refined set of rRNA-targeted oligonucleotide probes for the *in situ* detection and quantification of ammonia-oxidizing bacteria ***In prep***

Paper 3 JM. Kristensen, C. Singleton, F. Petriglieri, L. Clegg, PH. Nielsen. Acidobacteria in activated sludge: diversity, distribution, and genomics ***In prep***

Paper 4 JM. Kristensen, M. Nierychlo, M. Albertsen, PH. Nielsen. Bacteria from the genus *Arcobacter* are abundant in effluent from wastewater treatment plants ***Submitted***

Papers not included in this thesis

Paper 5 SJ. McIlroy, RH. Kirkegaard, B. McIlroy, M. Nierychlo, **JM Kristensen**, SM. Karst, M. Albertsen, PH. Nielsen. MiDAS 2.0: an ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups. Database. 2017. ***Published***

Paper 6 RH. Kirkegaard, SJ. McIlroy, **JM. Kristensen**, M. Nierychlo, SM Karst, MS. Dueholm, M. Albertsen, PH, Nielsen. The impact of immigration on microbial community composition in full-scale anaerobic digesters. Scientific Reports. 2017. ***Published***

Paper 7 SJ. McIlroy, CA. Onetto, B. McIlroy, FA. Herbst, MS. Dueholm, RH. Kirkegaard, E. Fernando, SM. Karst, M. Nierychlo, **JM. Kristensen**, KL. Eales, PR. Grbin, R. Wimmer, PH. Nielsen. Genomic and in Situ Analyses Reveal the Micropruina spp. as Abundant Fermentative Glycogen Accumulating Organisms in Enhanced Biological Phosphorus Removal Systems. Frontiers in Microbiology. 2018. ***Published***

Paper 8 G. Dottorini, TY. Michaelsen, KS. Andersen, **JM. Kristensen**, DS. Wágner, M. Nierychlo, PH. Nielsen. Immigration determines microbial community assembly in water resource recovery systems. ***In prep***

CHAPTER 1. INTRODUCTION

1.1. BIOLOGICAL WASTEWATER TREATMENT AND MICROBIAL PROCESSES

Microbes are the main drivers of many natural processes and global cycles. They are responsible for converting biomass to carbon dioxide and methane, turning over nitrogen compounds and many other processes that are vital for life on earth. The versatility of microbial conversion has been exploited in engineered systems and produces many important resources. An example is the wastewater treatment process which worldwide is carried out by the activated sludge process, a process where sewage is treated exploiting the natural metabolism of microorganisms and their floc forming abilities to clean the wastewater for nutrients such as carbon (C), nitrogen (N), phosphorous (P), micropollutants and pathogens. The wastewater treatment industry is the biggest biotechnological industry where the main workhorses are the bacteria[1]. The activated sludge treatment facilities have historically been operated for more than 100 years as a more or less "black box" system. Slowly biological process models have been inferred from the knowledge of very few organisms in pure culture divided based on their functional properties[2]. As an example the nitrifiers, *Nitrosomonas* genus discovered by Sergei Winogradsky in 1892[3]. In recent years advanced wastewater treatment plants now have a vital role in solving environmental challenges, ensuring human health protection from dangerous compounds and organisms and by reducing nutrient and pathogen content in the outlet[1].

A well-functioning wastewater treatment is increasingly important as a higher urbanization together with an increased human population is experienced globally. The rapid population growth comes with a risk of resource scarcity and has led to increased awareness of the need for sustainability and circular economy[4]. To meet the growing needs for resources and the production of waste streams in a sustainable way it is important to develop a more efficient use and recovery of existing resources[5]. Materials that were once considered waste and ended up in landfills or were incinerated are now increasingly being considered valuable resources. This is true for wastewater streams as well with a transition in mindset from costly wastewater treatment plants to resource recovery facilities, producing bioenergy, clean water, fertilizer etc[6]. In recent years wastewater treatment plants are increasingly referred to as resource recovery facilities. This change is due to the many possibilities derived from biological processes for recovery of certain substrates such as phosphorus, which can be utilized as e.g. fertilizer and the production of methane that can be used for bioenergy[7,8]. In a biological wastewater treatment plant **(Figure 1)**, the wastewater enters the biological treatment area with alternating conditions optimized for nutrient removal, thereafter it goes through a clarification tank where the biomass sludge is settled and removed as surplus sludge. The treated wastewater is after that discharged to the environment.

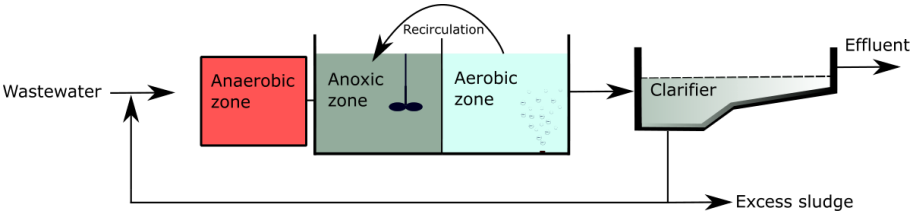

Figure 1 Schematic drawing of a wastewater treatment plant, inspired from Henze et al 2002[9].

Knowledge about the nitrogen cycle has been utilized for biological N- removal in wastewater treatment for many years to ensure a sufficient conversion of nitrogen compounds to dinitrogen gas and thus meet the end thresholds for treated wastewater. This is the total conversion from fixed nitrogen (ammonia, nitrate and nitrite) that if released to the environment can cause eutrophication, to dinitrogen gas that is released to the atmosphere[10]. These processes, termed nitrification, have for many years been thought of as a two-step process that happens in the aerobic zone of the activated sludge tank and involved two different groups of organisms[9]. First the ammonia oxidizing bacteria (AOBs) convert ammonia to nitrite, and then the nitrite oxidizing bacteria (NOBs) convert the nitrite to nitrate. However, recently bacteria with the ability to perform complete ammonia oxidation in a single step was discovered and termed comammox bacteria[11,12] (CMXs). Following nitrification the nitrite and nitrate is removed in a process called denitrification, where facultative-anaerobic heterotrophic organisms use nitrate and/or nitrite as electron acceptors in a sequential reduction of nitrate into dinitrogen gas via several intermediates, this process happens in the anoxic zone of the activate sludge process[13]. Not all denitrifiers are able to converting nitrate all to nitrogen gas and thus this process can be incomplete, leaving compounds as nitric oxide NO or nitrous oxide (N_2O)[14]. Another catabolic process producing dinitrogen gas is the anaerobic ammonia oxidation (anammox) performed by special anammox bacteria. In the anammox process ammonia is converted into dinitrogen gas under anoxic conditions, using nitrite as electron acceptor[15].

Another important process in biological wastewater treatment is the enhanced biological phosphorous removal (EBPR) process. The EBPR-process setup (**Figure 1**) is designed to promote conditions to enrich for polyphosphate accumulating organisms (PAOs) which are able to store phosphate as intracellular polyphosphate and then remove them with the P trapped inside. These organisms are thus key for the biological phosphorus removal from wastewater[16]. The process principle is to cycle biomass under aerobic and anaerobic conditions. First under anaerobic conditions the PAOs gain a competitive advantage compared to other growing organisms, as they can take up volatile fatty acids and store them as carbon polymers the poly-β-hydroxyalkanoates (PHA). Under the following aerobic conditions, the PAOs are then able to use the stored compounds for rapid biomass growth and uptake of phosphate for excess storage as polyphosphate, the removal of P from the wastewater is then done by removal of the excess sludge with high polyphosphate content. This is how

the model PAO, *Accumulibacter,* is functioning. However, today it is known that also alternative PAOs exist, such as genus *Tetrasphaera*[17], that does not store PHA or cycle glycogen the same way, or organisms which can use nitrate or nitrite as electron acceptors instead of oxygen to perform simultaneously denitrification and polyphosphate accumulation[18,19]. Glycogen accumulating organisms (GAOs) do not cycle polyphosphate but compete with the PAOs for VFA and stores them as PHA under anaerobic conditions, under aerobic conditions they accumulate glycogen instead of polyphosphate[16].

These biological treatment and resource recovery processes are carried out by the microbial community within the activated sludge. Despite a huge effort through many years, the members of the microbial community in the sludge are to some extent still unknown[20]. To truly utilize the full potential of the wastewater treatment systems as well as other microbial processes the central questions to be answered are "Who is there?", "What can they do?", and "What are they doing?". Collecting knowledge and retrieving information about the members of the microbial community within the activated sludge is of utmost importance to gain a holistic understanding. An improved understanding can facilitate the design of new processes[21], biological monitoring, and control strategies.

1.2. CHARACTERIZING MICROBIAL COMMUNITIES IN ACTIVATED SLUDGE

Methods to identify and characterize functions of a bacteria have historically been culture-based, requiring isolation and culturing in the laboratory followed by various growth studies and biochemical profiling to work. Most current knowledge has been obtained using these slow and extremely time-consuming culture-based methods[22]. The study of pure cultures has great value but does not necessarily represent the right organisms or how the bacteria would behave *in situ* due to different *in vivo* properties and artificial growth conditions in a controlled laboratory environment. Furthermore, most organisms are not easy to grow in pure culture[23–25]. An early *in situ* method that has been applied to identify microorganisms has been by light microscopy analyzing their morphology and classifying them into morphotypes, but this has been found to be inadequate[26]. To gain knowledge of the remaining unknown bacteria the implementation of culture-independent methods is required.

Many modern culture-independent techniques for identification of bacteria rely on the small subunit ribosomal RNA (SSU rRNA) gene, which is a good phylogenetic marker gene because it represents a vital function for all organisms, protein synthesis, and contains both conserved and variable regions[27]. The SSU rRNA gene is still the most commonly used marker gene in microbial ecology and widely used in global microbial diversity studies such as the human microbiome project[28] and the Earth microbiome project[29]. The SSU rRNA gene is roughly 1500 bp long and codes for a piece of RNA that is an integrated part of the ribosome, and the alternating sequence

of variable and conserved regions (**Figure 2**) show that some regions are more important for the function of the ribosome than others. Using the SSU rRNA gene for identification, 16S for bacteria and archaea and 18S for eukaryotes, was first used by Woese and Fox[30] when their phylogenetic analysis in 1977 first defined the structure of the tree of life.

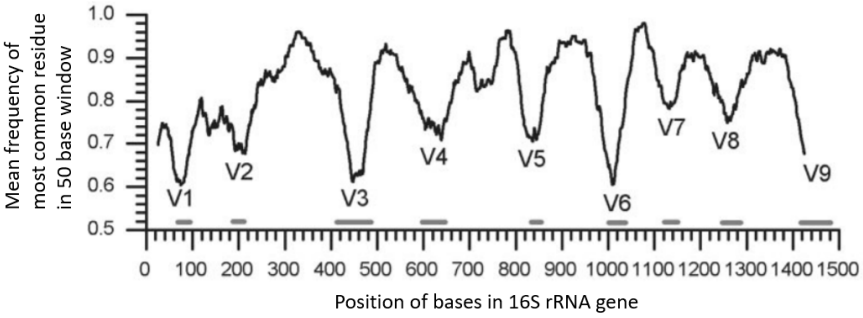

Figure 2 Distribution of conserved and variable nucleotides across the 16S rRNA gene from all known bacterial rRNA sequences (adapted from Ashelford et. al 2005[27]) the x-axis shows position of the base pair and the y-axis show the mean frequency of the most common bases at a specific position for all sequences on average.

Historically the most widely used culture-independent methods to obtain long 16S sequences has been the full cycle rRNA approach[25], which has relied on a direct DNA extraction followed by a PCR amplification step, cloning and DNA sequencing before a comparative Fluorescence *In Situ* Hybridization (FISH) analysis was used for verification.

FISH microscopy is a core method for studying bacteria in their natural environment[26]. This method uses synthetic DNA molecules called probes, which are typically 15-25 nucleotides long, and linked to a fluorescent dye. The DNA sequence is designed to be complementary to the region of the target rRNA sequence of the organism of interest. This way it is possible to mark only the cells with a matching rRNA sequence with a fluorescent dye, allowing identification and visualization under a microscope. The sequence specific labelling is necessary as many different bacteria cannot be distinguished based on other characteristics such as size and morphology. Visualization provides an insight that complements the high throughput oriented "omics" methods. It is possible to investigate information about co-localization, chemical composition of individual cells and more, which is lost in the "homogenize everything" approaches needed for most omics' methods. For this reason, FISH-based techniques and microscopy have not become obsolete even after many years but still play an important role in the study of microbial ecology. Many methods have been developed in combination with FISH microscopy to investigate *in situ* activity and

ecophysiology while linking the observed cell to an identity such as microautoradiography[31] and Raman microspectroscopy[32].

To study bacteria using FISH it is necessary to know which bacteria are present in a given environment and the sequence of their 16S rRNA genes in order to design a nucleotide probe that uniquely targets a given organism. However, the public databases of SSU sequences lack knowledge about bacteria from specific target ecosystems, such as wastewater treatment plants[33]. For design of FISH probes the lack of sequences is an issue that needs to be addressed in order to identify and visualize bacteria. The presence of sequences in the universal databases that are not found in the environment of interest can also be problematic, as it can make it impossible to identify unique sequences to use for organism specific probes. Even though the non-target organisms may require different conditions to grow and thus never will occur in the same ecosystem.

1.3. AIM

The overall objective of this PhD study was to gain detailed knowledge of physiology and ecology of novel or poorly described key species in activated sludge with the purpose of future application in process optimization and troubleshooting.

The main objectives are:

1. To develop an ecosystem-specific database of 16S rRNA gene sequences for activated sludge and utilize it to identify and gain knowledge of abundant novel bacteria or specific functional groups in wastewater treatment plants

2. To design species-specific FISH probes with multiple fluorochromes based on phylogenetic analyses of full-length 16S rRNA gene sequences

3. To gain a new understanding of spatial location and diversity of selected groups in the activated sludge by advanced microscopy and to explore their distribution and functions by genome analysis and Raman microspectroscopy

1.4. RELIABLE IDENTIFICATION AND VISUALIZATION USING FISH REQUIRE ROBUST REFERENCE DATABASES

The most comprehensive sequence databases for identification of bacteria are still based on the 16S rRNA gene and has been referred to as the "gold standard" in many cases for identification of bacteria[34]. Recently, the databases are being populated at an increasing rate with new branches of the tree of life due to development in sequencing techniques such as full-length 16S rRNA gene sequencing and metagenomics. As an example, the amount of 16S rRNA sequences deposited in the SILVA database has doubled in recent years[35] (**Figure 3**). However, we still have some way to go in the process of saturating the reference databases[36].

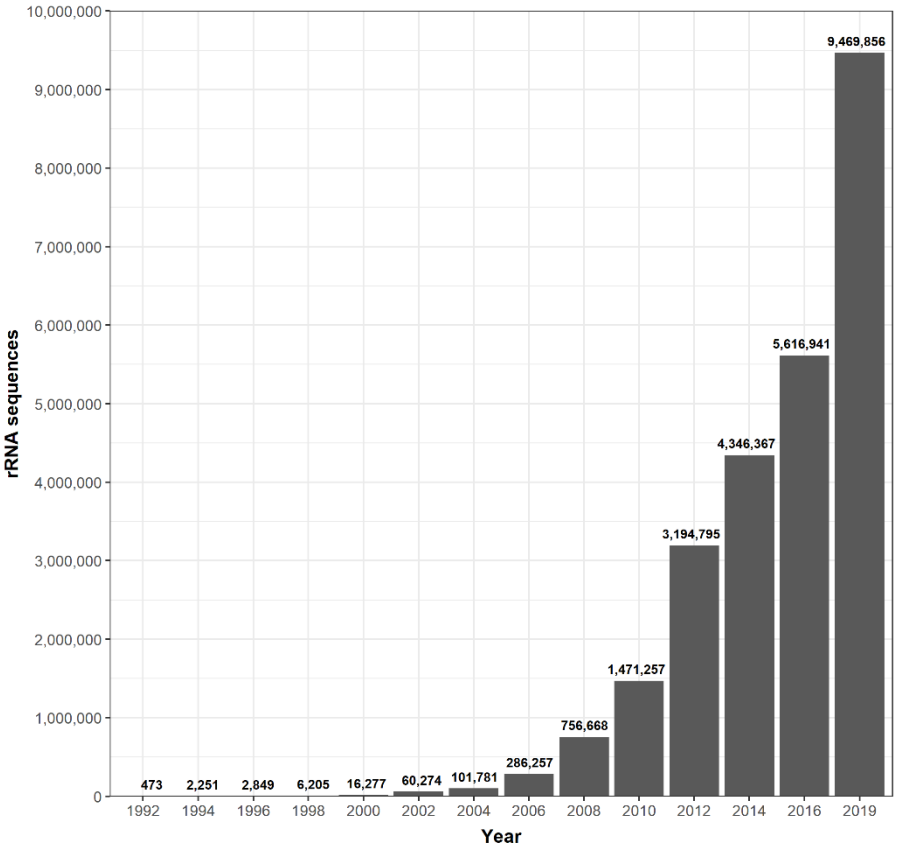

Figure 3 Growth of the SSU ribosomal databases from 1992 to 2019 (RDP II and SILVA), adapted from www.arb-silva.de

The current globally used 16S rRNA gene databases contain sequences for bacteria from many different environments and the availability of the 16S databases enable the development of single cell methods to study ecophysiology and activity *in situ*.

The rapid expansion of the 16S rRNA gene sequence databases also has some pitfalls when the databases are used for e.g. primer or probe design for FISH. The global databases have now been expanded to include millions of sequences from different environments, and the lack of metadata included in the databases makes it hard to differentiate them and to find their origin of environment and thus define the importance for the given environment that is being studied. The usage of different sequencing techniques, bioinformatic tools and quality thresholds in obtaining the sequences is biasing the outcome, and most importantly the use of the global databases is hampered by the lack of high-identity reference sequences specifically for wastewater treatment plants[20].

However, even if the databases are ever expanding, the most commonly used SILVA, Greengenes and RDP[37–39] still lack species specific information for many important bacteria abundant in activated sludge. With incomplete databases taxonomic classification relies on imperfect matches assigning names using classifier tools such as sintax[40] or RDP classifier[41]. Without a perfect match in the reference database it is often impossible to assign a species name and reliable identification at the lower taxonomic ranks. Without a perfect match or close relative in the reference database, the classification results in a higher taxonomic classification than it could have been based on the taxonomic resolution of the 16S rRNA gene. Due to these and other bioinformatical challenges it has often not been possible to gain reliable species identification of bacteria using the 16S gene.

To study community dynamics and functions in a wastewater treatment plant there is a need for species classification. as many functional traits are conserved at the genus or species level[42]. Furthermore, in some cases not all species or strains of the same genus have the same trait, such as pathogenicity[43] and antimicrobial resistance[44].

The SSU databases are used to design nucleotide probes for FISH. Designing a probe for FISH for a specific target require a robust and reliable database where high-quality reference sequences of the right bacteria are available. The key is then to define a target group and within the target group to find a unique sequence of 15-25 bp in the 16S rRNA gene that is shared by all the target members while different from the sequences in all other organisms within the database[45]. Designing FISH probes with large universal databases is becoming increasingly difficult due to the many new sequences, making it harder to find sites for differentiation of target groups, and especially for targeting lineage or species-specific bacteria.

The universal 16S databases contain millions of sequences but often without a specific match for the bacteria found *in situ*, due to the origin of the sequences. When designing a FISH-probe using a universal database as a backbone you can only target

what is already known and you must know all related bacteria of the whole community to be sure the probe only hits the target. As it is constructed now the current universal databases do not provide enough reference sequences to allow the design of probes and visualization of novelty for specific environments.

In order to ensure a reliable foundation for taxonomic classification, databases need to be comprehensive. Despite many years work in accumulating data, database completeness is still a major issue[36]. The widely used Silva database contains 10 million SSU sequences, which is still dwarfed by an estimate of 1 trillion different microbial species on the planet[46]. However, it is important to note that only 5 percent of the sequences are longer than 1,200 bp and thus include the full-length gene. This highlights that many of the sequences are incomplete and thus limiting the regions available for probe design for many species. An effect of having all organisms in the same database without proper metadata is that it becomes increasingly difficult to identify 15-25 bp that uniquely separate one organism from all others, making it impossible to design the perfect probe. If most of the closely related organisms never occur in your system of choice, then universal databases make it unnecessarily complicated to find a unique sequence and design a probe.

To mitigate some of these problems, system specific databases can be generated using the recent advances within 16S rRNA full-length sequencing by Karst et al. [35]. This method makes it feasible to move from using a database with largely unrelated sequences to generate a database with all the relevant sequences at scale. To link these novel sequences with the taxonomy, Dueholm *et al*. 2019 (see **Chapter 2**), developed a new method called Autotax for taxonomic assignment. This method was demonstrated for Danish wastewater treatment plants to develop a new ecosystem specific database for wastewater treatment and anaerobic digesters with the SILVA taxonomy as a backbone for taxonomy. 16S rRNA sequences has been obtained from environmental samples, using amplicon based methods as well as metagenome assembled genomes (MAGs), to generate ecosystem specific databases[37,47].

1.5. METHODS TO VISUALIZE MICROBIAL ECOLOGY

FISH has been used as one of the most commonly applied molecular methods for cultivation-independent study of microbial diversity and ecophysiology. The approach uses oligonucleotide probes which are covalently bound to a fluorescent dye, and then hybridized to the ribosomal RNA of specific targeted organisms, allowing detection of the fluorescent signal and *in situ* visualization of the microbial cells[25,48]. It has been used as a method for visualization for many years and there has been derived other methods both in combination with or derived from this principle, FISH is used for *in situ* abundance identification and quantification purposes[49,50], analyses of spatial distribution in biofilms and other samples with complex structures[51–54]. Methods have also been developed to enhance the fluorescence signal, such as catalyzed reporter deposition (CARD-FISH)[55]. Other methods combined with FISH allows cultivation-independent study of ecophysiology, examples of these methods are microautoradiography, nanometer scale secondary ion mass spectroscopy (NanoSIMS)[56,57] and Raman microspectroscopy[32,58], and biorthogonal noncanonical amino acid tagging method (BONCAT)[59], and methods to optimize probes for environmental bacteria has been developed to express the full length 16S rRNA gene in a cloned *E. coli* (Clone-FISH)[60]. It is important to choose the right FISH-methods for visualization to avoid method induced bias. The detection of some cells can be hindered using standard procedures destroying the structure of the bacterial flocs (Kristensen et al. **Chapter 5**). This may require modifications to preserve the structure during fixation as demonstrated in Lukumbuzya et al 2019[52].

Traditionally FISH methods have been visualized using mono-labeled probes with up to three different fluorescent dyes labelled at the 5'-ends. The fluorophores most commonly used for traditional FISH are the sulfoindocyanine dyes Cy3 and Cy5, and carboxyfluorescein-N-hydroxysuccinimide ester (FLUOS). One of the limitations for these dyes are their broad excitation and emission windows giving problems with excitation crosstalk and emission bleed-through of the fluorophores[61]. In recent years several methods have been developed to overcome these limitations, one of them is the multicolor DOPE-FISH method where signal is enhanced by attachment of an additional fluorophore to the oligonucleotide probe, the probe can then also be double-labelled with different dyes at their 5' and 3' ends[62,63]. With this method it is possible to visualize up to six FISH-probes simultaneously. Another method attaches up to four fluorophores (quadruple-labeled) to the probes, the method is called multi-labeled (MiL) FISH-probes and uses click-chemistry[64] to attach the fluorophores. Methods using different or combined colors uses color blending to identify different organisms. For this to work with multiple FISH-probes it requires a similar relative signal intensity for each dye, which by microscopy can be hard to achieve as it requires careful adjustments of camera exposure time and detector sensitivity[52]. Furthermore this can also be hard to achieve as different FISH-probes with the same dye have different signal properties due to differences in cellular ribosome content of the targets, and differences in permeabilization and hybridization efficiencies[65]. Another

multicolor FISH method is the CLASI-FISH method, where multiple mono-labeled probes in different dyes are targeting the same target organisms, and then the targets are identified during post processing of images using software that is unmixing the dye combinations. With this method it is possible to distinguish up to 15 organisms[61], but as with the DOPE and MiL-FISH methods it requires the binding affinities of the applied FISH-probes to be similar[63], which is rarely the case in complex samples with many different target groups.

One of the newest multicolor FISH methods, take advantage of the state-of-the-art confocal laser scanning microscopy with a super continuum white light laser (WLL) technology. This technology used a laser with tunable excitation wavelength at a large part of the visible light spectrum (470-670nm) and a high flexible window of emission detection in contrast to other point scan confocal lasers[52,66]. With the WLL technology, highly tunable detectors can be used in combination with alternative fluorescent dyes with different spectral properties (e.g. ATTO dyes) to differentiate signals in highly complex samples labelled with mono-labelled FISH-probes with little bleedthrough[52] (Lukumbuzya et al. (**Chapter 3**)). Another advantage of using these alternative dyes for FISH is that they often have a higher quantum yield, and are more photostable than traditional dyes[67]. The ATTO-dyes exhibit a long fluorescence decay lifetime (0.6-4.1 ns) which can be combined with time-gated detection systems to reduce interference from sources such as fluorophores with shorter lifetime, background and autofluorescence[68], ensuring a better performance in localization based imaging.

1.6. THE STATUS OF FISH PROBES FOR THE MOST ABUNDANT SPECIES FOUND IN GLOBAL AND DANISH ACTIVATED SLUDGE

Through the last 25 years FISH-probes has been designed to target the bacteria in the very diverse activated sludge. In recent years there have been published papers about the bacteria which are always present and active in the wastewater treatment plants, the so called activated sludge microbial core organisms[69] and the most important functional groups of organisms from wastewater treatment from all around the world[8,20,70]. The most striking about this new literature is that a lot of what has been discovered in the wastewater treatment plants has been novel organisms, and not many taxa are shared between the wastewater treatment plants of the world. In a study from the Global Water Microbiome Consortium where the microbial community of 269 wastewater treatment plants from 23 different countries on 6 continents were analyzed only 28 of 61.448 operational taxonomic units (OTUs) were shared among all of the wastewater treatment plants and on average these shared organisms accounted for only 12 % of the total read abundance in a sample[70]. For the taxa with a known functional trait, only the nitrifier *Nitrospira* and the PAO *Candidatus* Accumulibacter were present on this list of the global core microorganisms. Interestingly the model AOB *Nitrosomonas* was missing from the set of core organisms. Another survey of the microbial communities in Danish wastewater treatment plants (Nierychlo et al. 2019)[71] presented the 50 most abundant amplicon sequence variants (ASVs) of the Danish activated sludge. On this list *Tetrasphaera* was the most abundant PAO, not *Ca.* Accumulibacter. This indicates that also the functional taxa such as the AOB nitrifiers and PAOs are highly specialized to every wastewater treatment plant.

For the 28 OTU's found in the study by the Global Water Microbiome Consortium, only 13 of them had a genus classification which show the shared core consist of many novel taxa. In the Danish study similar to the Global Water Microbiome Consortium, there are also found a lot of novel bacteria as only 31 of 50 mentioned species had a genus classification. How similar these novel organisms are and how they are targeted by FISH probes is not possible to directly compare as different bioinformatical tools are used for data processing which is leading to differences in classification. Furthermore, these novel taxa needs to be assessed and defined in combination with other methods such as metagenomic studies to recover a reference sequence for them first, because OTU's and ASV's do not contain enough evolutionary information to ensure a reliable classification[22,72]. Therefore, the focus in this further analysis will be of the known abundant genera found in these two studies.

19 of the 30 most abundant known genera found either the global study or in Danish activated sludge have existing genera specific FISH-probes (**Table 1**).

Table 1: The 30 genera most often found from global and Danish wastewater treatment plants, the publication year of the FISH probes associated to their genera, and their function in the wastewater treatment plants.

Genus	FISH-Probe available (Year)	Role in wastewater treatment	Abundant in:
***Ca.* Accumulibacter**	2000[73]	PAO	Global/DK
Rhodoferax	2016[74]	Denitrification	Global/DK
Dechloromonas	2005[75], 2016[74]	Denitrification and putative PAO	Global/DK
Nitrospira	2001[76], 2006[77]	Nitrification	Global/DK
Dokdonella	No	Unknown	Global
Zoogloea	1995[78], 2008[79]	Denitrification	Global
Haliangium	2009[80], 2016[74]	Denitrification	Global
Arcobacter	1995[81], 1997[82]	Fermentation	Global
Cloacibacterium	No	Fermentation	Global
Turneriella	No	Unknown	Global
Zymomonas	No	Fermentation[83]	Global
***Ca.* Accumulimonas**	2012[84]	Putative PAO	Global
Acinetobacter	1994[85]	Abundant in influent	Global
Tetrasphaera	2005[86], 2019[33]	PAO	DK
Trichococcus	2001[87]	Fermentation	DK
***Ca.* Microthrix**	1997[88]	Filamentous	DK
Romboutsia	No	Fermentation[89]	DK
***Ca.* Villigracilis**	2019[90]	Fermentation	DK
Rhodobacter	1999[91]	Unknown	DK
Acidovorax	1996[92], 1999[93]	Denitrification	DK
Bradyrizobium	2018[94]	Fermentation	DK
Simplicispira	No	Unknown	DK
Lautropia	No	Unknown	DK
Ideonella	No	Unknown	DK
Sulfuritalea	2016[74]	Denitrification	DK
Clostridium_sensu_stricto_1	No	Unknown	DK
Propionicimonas	2012[95]	Fermentation	DK
***Ca.* Sarcinithrix**	2011[96]	Fermentation	DK
Ferruginibacter	No	Unknown	DK
Fodinicola	No	Unknown	DK

For 21 of the 30 most abundant genera their main functions in wastewater treatment have been described, this leaves 9 genera where the information about them is only sparsely known from either pure culture studies or other environments than wastewater and thus not necessary directly reflecting their role therein.

Of the 30 most often found known genera in global and Danish wastewater treatment plants, 11 of them have not been investigated by FISH and no available probes target these genera specifically. Of the 19 genera, which are targeted by one or more FISH-probes only seven of them (*Rhodoferax*, *Dechloromonas*, *Haliangium, Tetrasphaera, Bradyrizobium,* Candidatus *Villigracilis* and *Sulfuritalea*) have been designed or reevaluated in recent years (2016 – 2019) and many of the known 16S rRNA gene probes are potentially outdated as more sequences are now available. Many of the available probes are from the 1990's and are designed before the global databases were expanded and should therefore be reevaluated to be sure they still target the right groups within specific wastewater treatment plants especially when the specialized functional taxa are different.

1.7. A REVISION OF OLD FISH PROBES IS NESSESARY DUE TO ADVANCES IN DATABASES

The updated databases do not only give the opportunity to design new probes and visualize previously unknown bacteria, they also enable revision of past FISH probes. Many of the known FISH probes that target organism are from before the expansions of the 16S databases and therefore and should therefore be revisited.

A general probe for an important group of nitrifiers, the ammonia oxidizing bacteria (AOB), where a mixture of probes from Wagner et al. 1995, Mobarry et al. 1996, Juretschko et al. 1998 and Adamczyk et al. 2003 [26] has been used to visualize and quantify the population of *Nitrosomonas* and *Nitrosospira* AOBs in wastewater treatment plants for many years. However, while the old AOB-mix[26] still targets the known AOBs with the growing databases it is now shown to also target many other organisms outside the target group with the most recent SILVA database (Lukumbuzya et al. **Chapter 3**). Another probe set which like the AOB-mix also target non-specific organisms is the PAO-mix targeting the PAO *Ca.* Accumulibacter but at the same time also targets the GAO *Propionivibrio*[97].
This is highly problematic as there is a risk of visualizing non-target organisms and thus producing a wrong link between identity and functional interpretation. Designing a new updated FISH-probe set can reveal the morphology of non-target organisms hit by the old probe set. The specificity for this popular probe set in wastewater treatment has not been evaluated *in situ* for many years as the oldest probe in this probe set is from 1995. It is therefore unknown how often poor specificity produces erroneous results when FISH is applied.

Another example is for *Tetrasphaera* where the old probes from Kong et al. 2005, with an evaluation of the existing FISH-probes, it was seen that the existing *Tetrasphaera* probes did not effectively target the sequences of the abundant species, (Dueholm et al. **Chapter 2**) and the most abundant species was not targeted by the available probes, also resulting in erroneous results.

1.8. SPECIES-SPECIFIC FISH PROBES ARE NEEDED FOR NEXT GENERATION EVALUATION

When FISH probes are being reevaluated or redesigned it is important to investigate if it is possible to get to species level resolution as many traits may not be conserved above species or strain level. An example of this is from Dueholm et al. 2019 (**Chapter 2**) where it was demonstrated that it was possible to design species level specific probes for *Tetrasphaera* species and thereby analyse 4 very distinct phylotypes among the most abundant species. Many unknown organisms will also need to be studied to gain a complete understanding of the wastewater treatment system. One of the methods to study them *in situ* could be to design a FISH probe for them based on the ecosystem specific 16S rRNA gene databases and combine the study with either Raman spectroscopy or metagenomic investigations (**Chapter 4**). For some organisms the 16S rRNA gene might not have enough diversity to allow for species specific FISH-probes to be designed as it is known for *Nitrospira*[98], *Cyanobacteria*[99] and Actinobacteria[100]. In these cases, the resolution might only allow design of genus or family level probes. In addition to a sequence limited resolution, FISH probe design also relies on having the variation of the 16S rRNA gene in a position that is structurally accessible by FISH probes for hybridisation[26]. However even when the 16S rRNA gene sequences have some limitations in relation to FISH-probe design and may not always be the ideal gene for designing species-specific FISH-probes, it could still be used to visualize the target groups.

1.9. CONCLUSION

Full-length 16S rRNA gene sequences has been used to generate an ecosystem specific database for activated sludge (**Chapter 2**). The database has been used to design new FISH probes and study novel organisms and their functions in activated sludge systems (**Chapter 3 and 4**). It has furthermore become possible to design species specific FISH probes within the limitations of the resolution of the 16S rRNA gene.

The database provides new opportunities for FISH probe design. It is now possible to both reevaluate the specificity of old FISH probes and if necessary, design new probe sets. Examination of the existing probes for the AOB genera *Nitrosomonas* and *Nitrosospira* (**Chapter 3**) revealed that they did hit outgroups. For this reason, new probes were designed, and the specificity of the probe set was evaluated *in situ* with multicolor FISH.

In combination with recent advances in confocal microscopy and FISH-methods it has then become possible to visualize probe sets with novel or uncharacterized organisms. The database was utilized in combination with FISH to investigate the potential role of bacteria belonging to the phylum Acidobacteria in wastewater treatment (**Chapter 4**). Metagenomic assembled genomes in combination with FISH-Raman revealed some to be putative PAOs.

The database was used to investigate the fate of pathogenic bacteria from the genus *Arcobacter* in wastewater treatment plants (**Chapter 5**). Cells from the genus *Arcobacter* were found to be only partially removed by the wastewater treatment process, suggesting cells from this genus to be loosely attached to the activated sludge floc and therefore staying in the water phase. The viability and pathogenicity were confirmed by isolation and genome sequencing.

On a global scale a lot of novelty still needs to be assessed and visualized as not many bacterial taxa are shared within wastewater treatment plants suggesting a lot of diversity that still needs to be uncovered.

Of the known genera abundant on a global and Danish scale only 19 of 30 genera had a specific FISH probe to target them. Only few of these probes have been designed recent enough to consider the massive growth of reference databases. A reevaluation of the old probe sets is thus necessary, as it was revealed that the old probe sets can inadvertently target different previously uncharacterized outgroups of bacteria in the activated sludge. However, the opposite can also be true as genus level probes can miss specific species found in the wastewater treatment plants as the communities here are highly specialized (**Chapter 2**).

1.10. PERSPECTIVES

The ecosystem specific 16S rRNA databases are a useful tool for designing and optimizing FISH-probes, here the database is based on the specific wastewater treatment microbial community and but the method itself could extended to any other environment, facilitating FISH probe design. Others have already made ecosystem specific databases for other environments like the oral microbiome[47] and the human gut[101], here the primary reason is also for a better identification, but it is also to reduce computational time by having less references to match sequences up against. Even though the 16S rRNA gene has potential for 4^{1500} variants, constraints on structure and function limits this in reality, making it impossible to distinguish all organisms with FISH probes using this gene alone. Organisms can share the same exact 16S rRNA sequence and carry out very different functions[102]. This effectively makes it impossible to design species specific probes based on the 16S rRNA sequence for these species. This highlights a need to move beyond single marker gene-based probes and harbor the full benefit of complete genomes, as a start the 23S rRNA gene could also be included in database. Other marker genes with better resolution for specific organisms could be explored. However, with this comes the challenge of generating enough signal as there is a big difference moving from targeting the many ribosomes per cell to a single gene or genome per cell. By including tags for the whole genome sequence, it would be truly possible to visualize on a species or strain if needed.

Another important factor besides making FISH more specific would be for FISH to be upscaled and automated to be make the transition to high throughput. Today it takes a lot of time and specialized staff to perform FISH and microscopy. Implementing and combining multicolor FISH using alternative dyes with automated microscopy could make the process more feasible. It could be important to upscale microscopy-based methods in the same way omics techniques have been enhanced in recent years. Furthermore, high throughput FISH will pave the way for upscaling more applied techniques such as qFISH. Another new method that is currently being implemented, especially in detection of larger cells in the medical field, is deep learning for automatic segmentation and recognition of different cell types. This method could in the future also be implemented in the field to create a fully automated high throughput FISH process.

LITERATURE

1. Nielsen, P. H., Saunders, A. M., Hansen, A. A., Larsen, P. & Nielsen, J. L. Microbial communities involved in enhanced biological phosphorus removal from wastewater - a model system in environmental biotechnology. *Curr. Opin. Biotechnol.* **23**, 452–459 (2012).

2. Nielsen, P. H. *et al.* A conceptual ecosystem model of microbial communities in enhanced biological phosphorus removal plants. *Water Res.* **44**, 5070–5088 (2010).

3. Winogradsky, S. Contributions a la morphologie des organismes de la nitrification. *Arch. Sci. Biol.* **1**, 88–137 (1892).

4. van Loosdrecht, M. C. M. & Brdjanovic, D. Water treatment. Anticipating the next century of wastewater treatment. *Science* **344**, 1452–3 (2014).

5. Trimmer, J. T., Miller, D. C. & Guest, J. S. Resource recovery from sanitation to enhance ecosystem services. *Nat. Sustain.* **2**, 681–690 (2019).

6. Nielsen, P. H. Microbial biotechnology and circular economy in wastewater treatment. *Microb. Biotechnol.* **10**, 1102–1105 (2017).

7. Nielsen, Per Halkjær, Seviour, R. J. *Microbial Ecology of Activated Sludge*. (IWA Publishing, 2010).

8. McIlroy, S. J. *et al.* MiDAS 2.0: an ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups. *Database (Oxford).* **2017**, (2017).

9. Henze, M., Harremoës, P., la Cour Jansen, J. & Arvin, E. *Wastewater treatment. Biological and chemical processes*. (Polyteknisk forlag, 2002).

10. Henze, M., Van Loosdrecht, M. C. M., Ekama, G. A. & Brdjanovic, D. *Biological Wastewater treatment, Principles, Modelling and Design*. (IWA Publishing, 2008).

11. Daims, H. *et al.* Complete nitrification by *Nitrospira* bacteria. *Nature* **528**, 504–509 (2016).

12. van Kessel, M. A. H. J. *et al.* Complete nitrification by a single microorganism. *Nature* **528**, 555–9 (2015).

13. Daims, H. & Wagner, M. The microbiology of nitrogen removal. in *Microbial*

ecology of activated sludge; Seviour, R. J.; Nielsen, P. H., 259–280 (IWA Publishing, 2010).

14. Zumft, W. G. Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* **61**, 533–616 (1997).

15. Jetten, M. S. M. *et al.* Biochemistry and molecular biology of anammox bacteria biochemistry and molecular biology of anammox bacteria. *Crit. Rev. Biochem. Mol. Biol.* **44**, 65–84 (2009).

16. Oehmen, A. *et al.* Advances in enhanced biological phosphorus removal: From micro to macro scale. *Water Res.* **41**, 2271–2300 (2007).

17. Kristiansen, R. *et al.* A metabolic model for members of the genus *Tetrasphaera* involved in enhanced biological phosphorus removal. *ISME J.* **7**, 543–554 (2013).

18. Kerrn-Jespersen, J. P. & Henze, M. Biological phosphorus uptake under anoxic and aerobic conditions. *Water Res.* **27**, 617–624 (1993).

19. Hu, Z. R., Wentzel, M. C. & Ekama, G. A. Anoxic growth of phosphate-accumulating organisms (PAOs) in biological nutrient removal activated sludge systems. *Water Res.* **36**, 4927–4937 (2002).

20. McIlroy, S. J. *et al.* MiDAS: The field guide to the microbes of activated sludge. *Database* **2015**, 1–8 (2015).

21. Mulder, A., van de Graaf, A. A., Robertson, L. A. & Kuenen, J. G. Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor. *FEMS Microbiol. Ecol.* **16**, 177–183 (1995).

22. Yarza, P. *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* **12**, 635–645 (2014).

23. Stewart, E. J. Growing unculturable bacteria. *J. Bacteriol.* **194**, 4151–4160 (2012).

24. Imachi, H. *et al.* Isolation of an archaeon at the prokaryote-eukaryote interface. *bioRxiv* (2019). doi:10.1101/726976

25. Amann, R. I., Ludwig, W. & Schleifer, K. H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**, 143–169 (1995).

26. Nielsen, P. H. *FISH Handbook for Biological Wastewater Treatment*. *Water Intelligence Online* **8**, (2009).

27. Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J. & Weightman, A. J. At Least 1 in 20 16S rRNA Sequence Records Currently Held in Public Repositories Is Estimated To Contain Substantial Anomalies. *Appl. Environ. Microbiol.* **71**, 7724–7736 (2005).

28. Turnbaugh, P. J. *et al.* The Human Microbiome Project. *Nature* **449**, 804–810 (2007).

29. Gilbert, J. A., Jansson, J. K. & Knight, R. The Earth Microbiome project: Successes and aspirations. *BMC Biol.* **12**, 1–4 (2014).

30. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. USA* **74**, 5088–5090 (1977).

31. Lee, N. *et al.* Combination of fluorescent in situ hybridization and microautoradiography - a new tool for structure-function analyses in microbial ecology. *Appl. Environ. Microbiol.* **65**, 1289–1297 (1999).

32. Fernando, E. Y. *et al.* Resolving the individual contribution of key microbial populations to enhanced biological phosphorus removal with Raman–FISH. *ISME J.* **13**, 1933–1946 (2019).

33. Dueholm, M. S. *et al.* Comprehensive ecosystem-specific 16S rRNA gene databases with automated taxonomy assignment (AutoTax) provide species-level resolution in microbial ecology. *bioRxiv* (2019). doi:10.1101/672873

34. Janda, J. M. & Abbott, S. L. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *J. Clin. Microbiol.* **45**, 2761–2764 (2007).

35. Karst, S. M. *et al.* Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. *Nat. Biotechnol.* **36**, 190–195 (2018).

36. Schloss, P. D., Girard, R. A., Martin, T., Edwards, J. & Thrash, J. C. Status of the archaeal and bacterial census: An update. *MBio* **7**, 1–10 (2016).

37. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, 590–596 (2013).

38. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene

database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).

39. Cole, J. R. *et al.* Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **42**, 633–642 (2014).

40. Edgar, R. SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. *bioRxiv* (2016). doi:10.1101/074161

41. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).

42. Martiny, J. B. H., Jones, S. E., Lennon, J. T. & Martiny, A. C. Microbiomes in light of traits: A phylogenetic perspective. *Science* **350**, aac9323 (2015).

43. Pupo, G. M., Karaolis, D. K. R., Lan, R. & Reeves, P. R. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. *Infect. Immun.* **65**, 2685–2692 (1997).

44. Chambers, H. F. & Deleo, F. R. Waves of Resistance: *Staphylococcus aureus* in the Antibiotic Era. *Nat. Rev. Microbiol.* **7**, 629–641 (2010).

45. Ludwig, W. *et al.* ARB : a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).

46. Locey, K. J. & Lennon, J. T. Scaling laws predict global microbial diversity. **2016**, 30–35 (2016).

47. Dewhirst, F. E. *et al.* The human oral microbiome. *J. Bacteriol.* **192**, 5002–5017 (2010).

48. DeLong, E. F., Wickham, G. S. & Pace, N. R. Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. *Science* **243**, 1360–3 (1989).

49. Wagner, M. *et al.* Combining Fluorescent in situ hybridization (FISH) with cultivation and mathematical modeling to study population structure and function of ammonia-oxidizing bacteria in activated sludge. *Water Sci. Technol.* **37**, 441–449 (1998).

50. Daims, H. & Wagner, M. Quantification of uncultured microorganisms by fluorescence microscopy and digital image analysis. *Appl. Microbiol.*

Biotechnol. **75**, 237–248 (2007).

51. Daims, H., Lücker, S. & Wagner, M. Daime, a Novel Image Analysis Program for Microbial Ecology and Biofilm Research. *Environ. Microbiol.* **8**, 200–213 (2006).

52. Lukumbuzya, M., Schmid, M., Pjevac, P. & Daims, H. A multicolor fluorescence in situ hybridization approach using an extended set of fluorophores to visualize microorganisms. *Front. Microbiol.* **10**, 1–13 (2019).

53. Almstrand, R., Daims, H., Persson, F., Sörensson, F. & Hermansson, M. New methods for analysis of spatial distribution and coaggregation of microbial populations in complex biofilms. *Appl. Environ. Microbiol.* **79**, 5978–5987 (2013).

54. Dolinšek, J., Lagkouvardos, I., Wanek, W., Wagner, M. & Daims, H. Interactions of nitrifying bacteria and heterotrophs: Identification of a Micavibrio-like putative predator of *Nitrospira* spp. *Appl. Environ. Microbiol.* **79**, 2027–2037 (2013).

55. Pernthaler, A., Pernthaler, J. & Amann, R. Fluorescence In Situ Hybridization and Catalyzed Reporter Deposition for the Identification of Marine Bacteria. *Appl. Environ. Microbiol.* **68**, 3094–3101 (2002).

56. Musat, N., Foster, R., Vagner, T., Adam, B. & Kuypers, M. M. M. Detecting metabolic activities in single cells, with emphasis on nanoSIMS. *FEMS Microbiol. Rev.* **36**, 486–511 (2012).

57. Berry, D. *et al.* Host-compound foraging by intestinal microbiota revealed by single-cell stable isotope probing. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 4720–4725 (2013).

58. Huang, W. E. *et al.* Raman-FISH: combining stable-isotope Raman spectroscopy and fluorescence in situ hybridization for the single cell analysis of identity and function. *Soc. Appl. Microbiol.* **9**, 1878–1889 (2007).

59. Hatzenpichler, R. *et al.* In situ visualization of newly synthesized proteins in environmental microbes using amino acid tagging and click chemistry. *Environ. Microbiol.* **16**, 2568–2590 (2014).

60. Schramm, A., Fuchs, B. M., Nielsen, J. L., Tonolla, M. & Stahl, D. A. Fluorescence in situ hybridization of 16S rRNA gene clones (Clone-FISH) for probe validation and screening of clone libraries. *Environ. Microbiol.* **4**, 713–720 (2002).

61. Valm, A. M., Mark Welch, J. L. & Borisy, G. G. CLASI-FISH: Principles of Combinatorial Labeling and Spectral Imaging. *Syst. Appl. Microbiol.* **23**, 1–7 (2012).

62. Stoecker, K., Dorninger, C., Daims, H. & Wagner, M. Double labeling of oligonucleotide probes for fluorescence in situ hybridization (DOPE-FISH) improves signal intensity and increases rRNA accessibility. *Appl. Environ. Microbiol.* **76**, 922–926 (2010).

63. Behnam, F., Vilcinskas, A., Wagner, M. & Stoeckerb, K. A straightforward DOPE (double labeling of oligonucleotide probes)-FISH (fluorescence in situ hybridization) method for simultaneous multicolor detection of six microbial populations. *Appl. Environ. Microbiol.* **78**, 5138–5142 (2012).

64. Schimak, M. P. *et al.* MiL-FISH: Multilabeled oligonucleotides for fluorescence in situ hybridization improve visualization of bacterial cells. *Appl. Environ. Microbiol.* **82**, 62–70 (2016).

65. Wagner, M., Horn, M. & Daims, H. Fluorescence in situ hybridisation for the identification and characterisation of prokaryotes. *Curr. Opin. Microbiol.* **6**, 302–309 (2003).

66. McConnell, G. Confocal laser scanning fluorescence microscopy with a visible continuum source. *Opt. Express* **12**, 2844 (2004).

67. Dempsey, G. T., Vaughan, J. C., Chen, K. H. & Bates, M. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nat. Methods* **8**, 1027–1036 (2011).

68. Rich, R. M. *et al.* Elimination of autofluorescence background from fluorescence tissue images by use of time-gated detection and the AzaDiOxaTriAngulenium (ADOTA) fluorophore. *Anal. Bioanal. Chem.* **405**, 2065–2075 (2013).

69. Saunders, A. M., Albertsen, M., Vollesen, J. & Nielsen, P. H. The activated sludge ecosystem contains a core community of abundant organisms. *ISME J.* **10**, 1–10 (2015).

70. Wu, L. *et al.* Global diversity and biogeography of bacterial communities in wastewater treatment plants. *Nat. Microbiol.* **4**, 1183–1195 (2019).

71. Nierychlo, M. *et al.* Species-level microbiome composition of activated sludge - introducing the MiDAS 3 ecosystem-specific reference database and taxonomy. *bioRxiv* (2019). doi:10.1017/CBO9781107415324.004

72. Edgar, R. C. Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. *PeerJ* **2018**, 1–29 (2018).

73. Crocetti, G. R. *et al.* Identification of polyphosphate-accumulating organisms and design of 16S rRNA-directed probes for their detection and quantitation. *Appl. Environ. Microbiol.* **66**, 1175–1182 (2000).

74. McIlroy, S. J. *et al.* Identification of active denitrifiers in full-scale nutrient removal wastewater treatment systems. *Environ. Microbiol.* **18**, 50–64 (2016).

75. Ginige, M. P., Keller, J. & Blackall, L. L. Investigation of an acetate-fed denitrifying microbial community by stable isotope probing, full-cycle rRNA analysis, and fluorescent in situ hybridization-microautoradiography. *Appl. Environ. Microbiol.* **71**, 8683–8691 (2005).

76. Daims, H., Nielsen, J. L., Nielsen, P. H., Schleifer, K.-H. & Wagner, M. In Situ Characterization of *Nitrospira*-Like Nitrite-Oxidizing Bacteria Active in Wastewater Treatment Plants. *Appl. Environ. Microbiol.* **67**, 5273–5284 (2001).

77. Maixner, F. *et al.* Nitrite concentration influences the population structure of *Nitrospira*-like bacteria. *Environ. Microbiol.* **8**, 1487–1495 (2006).

78. Rossello-Mora, R. A., Wagner, M., Amann, R. & Schleifer, K. H. The abundance of *Zoogloea ramigera* in sewage treatment plants. *Appl. Environ. Microbiol.* **61**, 702–707 (1995).

79. Oshiki, M., Satoh, H., Mino, T. & Onuki, M. PHA-accumulating microorganisms in full-scale wastewater treatment plants. *Water Sci. Technol.* **58**, 13–20 (2008).

80. Kyselková, M. *et al.* Comparison of rhizobacterial community composition in soil suppressive or conducive to tobacco black root rot disease. *ISME J.* **3**, 1127–1138 (2009).

81. Wesley, I. V *et al. Arcobacter*-Specific and *Arcobacter butzleri*-Specific 16S rRNA-Based DNA Probes. *J. Clin. Microbiol.* **33**, 1691–1698 (1995).

82. Snaidr, J., Amann, R., Huber, I. & Ludwig, W. Phylogenetic analysis and in situ identification of bacteria in activated sludge. *Appl. Environ. Microbiol.* **63**, 2884–2896 (1997).

83. Yang, S. *et al. Zymomonas mobilis* as a model system for production of

biofuels and biochemicals. *Microb. Biotechnol.* **9**, 699–717 (2016).

84. Nguyen, H. T. T., Nielsen, J. L. & Nielsen, P. H. '*Candidatus* Halomonas phosphatis', a novel polyphosphate-accumulating organism in full-scale enhanced biological phosphorus removal plants. *Environ. Microbiol.* **14**, 2826–2837 (2012).

85. Wagner, M. *et al.* Development of an rRNA-targeted oligonucleotide probe specific for the genus *Acinetobacter* and its application for in situ monitoring in activated sludge. *Appl. Environ. Microbiol.* **60**, 792–800 (1994).

86. Kong, Y., Nielsen, J. L. & Nielsen, P. H. Identity and ecophysiology of uncultured actinobacterial polyphosphate-accumulating organisms in full-scale enhanced biological phosphorus removal plants. *Appl. Environ. Microbiol.* **71**, 4076–4085 (2005).

87. Liu, J. R. & Seviour, R. J. Design and application of oligonucleotide probes for fluorescent in situ identification of the filamentous bacterial morphotype Nostocoida limicola in activated sludge. *Environ. Microbiol.* **3**, 551–560 (2001).

88. Erhart, R., Bradford, D., Seviour, R. J., Amann, R. & Blackall, L. L. Development and use of fluorescent in situ hybridization probes for the detection and identification of 'Microthrix parvicella' in activated sludge. *Syst. Appl. Microbiol.* **20**, 310–318 (1997).

89. Gerritsen, J. *et al.* Characterization of *Romboutsia ilealis* gen. nov., sp. nov., isolated from the gastro-intestinal tract of a rat, and proposal for the reclassification of five closely related members of the genus Clostridium into the genera *Romboutsia* gen. no. *Int. J. Syst. Evol. Microbiol.* **64**, 1600–1616 (2014).

90. Nierychlo, M. *et al.* The morphology and metabolic potential of the Chloroflexi in full-scale activated sludge wastewater treatment plants. *FEMS Microbiol. Ecol.* **95**, fiy228 (2019).

91. Giuliano, L., De Domenico, M., De Domenico, E., Höfle, M. G. & Yakimov, M. M. Identification of Culturable Oligotrophic Bacteria within Naturally Occurring Bacterioplankton Communities of the Ligurian Sea by 16S rRNA Sequencing and Probing. *Microb. Ecol.* **37**, 77–85 (1999).

92. Amann, R. *et al.* rRNA-targeted oligonucleotide probes for the identification of genuine and former pseudomonads. *Syst. Appl. Microbiol.* **19**, 501–509 (1996).

93. Schulze, R. *et al.* Genotypic diversity of *Acidovorax* strains isolated from activated sludge and description of *Acidovorax defluvii* sp. nov. *Syst. Appl. Microbiol.* **22**, 205–214 (1999).

94. Tominski, C. *et al.* Insights into Carbon Metabolism Provided by Fluorescence In Situ Hybridization-Secondary Ion Mass Spectrometry Imaging of an Autotrophic, Nitrate-Reducing, Fe(II)-Oxidizing Enrichment Culture. *Appl. Environ. Microbiol.* **84**, 1–19 (2018).

95. Nielsen, J. L., Nguyen, H., Meyer, R. L. & Nielsen, P. H. Identification of glucose-fermenting bacteria in a full-scale enhanced biological phosphorus removal plant by stable isotope probing. *Microbiol. (United Kingdom)* **158**, 1818–1825 (2012).

96. Speirs, L. B. M., McIlroy, S. J., Petrovski, S. & Seviour, R. J. The activated sludge bulking filament Eikelboom morphotype 0914 is a member of the Chloroflexi. *Environ. Microbiol. Rep.* **3**, 159–165 (2011).

97. Albertsen, M., McIlroy, S. J., Stokholm-Bjerregaard, M., Karst, S. M. & Nielsen, P. H. '*Candidatus* Propionivibrio aalborgensis': A novel glycogen accumulating organism abundant in full-scale enhanced biological phosphorus removal plants. *Front. Microbiol.* **7**, 1–17 (2016).

98. Pester, M. *et al.* NxrB encoding the beta subunit of nitrite oxidoreductase as functional and phylogenetic marker for nitrite-oxidizing *Nitrospira*. *Environ. Microbiol.* **16**, 3055–3071 (2014).

99. Moore, L. R., Rocap, G. & Chisholm, S. W. Physiology and molecular phytogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**, 464–467 (1998).

100. Nouioui, I. *et al.* Genome-based taxonomic classification of the phylum Actinobacteria. *Front. Microbiol.* **9**, 1–119 (2018).

101. Ritari, J., Salojärvi, J., Lahti, L. & de Vos, W. M. Improved taxonomic assignment of human intestinal 16S rRNA sequences by a dedicated reference database. *BMC Genomics* **16**, 1–10 (2015).

102. Antony-Babu, S. *et al.* Multiple *Streptomyces* species with distinct secondary metabolomes have identical 16S rRNA gene sequences. *Sci. Rep.* **7**, 1–8 (2017).

CHAPTER 2. COMPREHENSIVE ECOSYSTEM-SPECIFIC 16S RRNA GENE DATABASES WITH AUTOMATED TAXONOMY ASSIGNMENT (AUTOTAX) PROVIDE SPECIES-LEVEL RESOLUTION IN MICROBIAL ECOLOGY

MS. Dueholm, KS. Andersen, F. Petriglieri, SJ. McIlroy, M. Nierychlo, J. Petersen, **JM. Kristensen**, E. Yashiro, SM. Karst, M. Albertsen, and PH. Nielsen. Comprehensive ecosystem-specific 16S rRNA gene databases with automated taxonomy assignment (AutoTax) provide species-level resolution in microbial ecology

Preprint on BioRxiv

CHAPTER 3. A REFINED SET OF RRNA-TARGETED OLIGONUCLEOTIDE PROBES FOR THE IN SITU DETECTION AND QUANTIFICATION OF AMMONIA-OXIDIZING BACTERIA

M. Lukumbuzya, **JM. Kristensen**, P. Pjevac, K. Kitzinger, A. Pommerening-Röser, PH. Nielsen, H. Daims, M. Wagner. A refined set of rRNA-targeted oligonucleotide probes for the in situ detection and quantification of ammonia-oxidizing bacteria

In preparation

CHAPTER 4. ACIDOBACTERIA IN ACTIVATED SLUDGE: DIVERSITY, DISTRIBUTION, AND GENOMICS

JM. Kristensen, C. Singleton, F. Petriglieri, L. Clegg, PH. Nielsen. Acidobacteria in activated sludge: diversity, distribution, and genomics

In preparation

CHAPTER 5. BACTERIA FROM THE GENUS *ARCOBACTER* ARE ABUNDANT IN EFFLUENT FROM WASTEWATER TREATMENT PLANTS

JM. Kristensen, M. Nierychlo, M. Albertsen, PH. Nielsen. Bacteria from the genus *Arcobacter* are abundant in effluent from wastewater treatment plants

Submitted to Applied and Environmental Microbiology

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-576-5 AALBORG UNIVERSITY PRESS